Journal of Biomedical Informatics 52 (2014) 1–3

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Guest Editorial

Advancing methodologies in Clinical Research Informatics (CRI): Foundational work for a maturing field

Improvements in human health and wellness rely on the efficient and effective conduct of clinical research. As defined by the National Institutes of Health, *clinical research* encompasses a range of activities including patient-oriented research, epidemiology and behavioral studies, and outcomes research and health services research [1]. Conducting clinical research is a complex and information-intensive endeavor, often involving multiple actors, workflows, and processes, as well as a range of data management, analytics, and socio-technical factors. Such challenges have only increased in recent years owing to a shift toward more integrative, translational, and multi-institutional research activities. Given the growing need to accelerate clinical research and the myriad information-related challenges inherent in doing so, the emergent informatics sub-domain of clinical research informatics (CRI) has experienced a steady growth of activity [2–10].

Over the past decade, with funding from institutes and agencies charged with accelerating biomedical science [3,9,11,12], increasing numbers of investigations have led to significant improvements in the quality and efficiency of clinical research [6,13–15]. Such efforts have yielded advances in research systems and data management solutions related to the range of clinical research from clinical trials to comparative effectiveness and health services research [16–18]. In addition to these initiatives focused directly on improving applied CRI practice, investments and programs designed to increase adoption and "meaningful use" of Health IT (HIT) for clinical practice have also contributed indirectly to advances in CRI. Indeed, the increasing adoption and use of EHRs and improvements in standards-based data sharing hold great promise not only for improving healthcare but also for accelerating research via data re-use and enabling an "evidence generating medicine" practice paradigm to help create a learning health system [19,20].

As CRI has matured in recent years, the preponderance of CRI studies in the peer-reviewed literature have reported on application-specific tool development and the verification or validation of such tools given targeted use-cases. While these studies are extremely important to the field, advancing the science of CRI also requires studies and reports of extensible and empirically validated informatics methodologies that address fundamental clinical research relevant information needs. Through the development and dissemination of broad, integrative and cross-cutting theories and methodologies, the CRI community can advance the field and develop much-needed solutions to as yet unsolved data-, information-, and knowledge-management challenges in clinical research.

Given this need and the relative paucity of reports of CRI methods in the literature to date, we disseminated a call last year for papers focused on methodological issues in CRI. This special issue is the result of the CRI community's response to that call and the rigorous peer-reviewed process that followed. While many excellent papers were submitted, not all could be accepted. As such, this issue provides but a snapshot of the excellent work being done to advance our field via the development and study of CRI methods.

While the range of methodological issues to address in CRI are as varied as the informatics and research processes they are meant to address, early work in this area appears to be focused on a number of pressing, important and foundational issues. As the articles in this special issue reflect, some of the work in this space can be categorized as belonging to one of several categories. These include ongoing work on CRI methods focused on enabling:

- The secure, reliable, and reproducible capture, collection, and re-use of data for research purposes.
- Standards for consistently modeling and devising informatics solutions to clinical research information needs.
- Approaches to accelerate science appropriately via informatics while balancing the very real and critical regulatory and ethical issues inherent in such activities.
- Better understanding, visualization, and facilitation of research collaboration networks and team science via informatics approaches.
- Improved user experience and CRI solution development through better understanding of research users' information needs.
- The integration of clinical practice and research activities in order to leverage our health system for evidence generation.

On the topic of data collection as a critical first step to enable data re-use, frameworks such as the one for medication data collection across research networks described by Richesson, inform rational and standardized approaches to such activities. Once collected, data storage and retrieval become the next focus of activity for advanced CRI methods to enable data re-use for research [21]. It is therefore not surprising that issues related to data storage and retrieval to enable research are addressed by several of the reports herein, including: the work of Cimino et al. on the design considerations for advanced information systems for data storage and retrieval [22]; the description of frameworks to guide institutional

http://dx.doi.org/10.1016/j.jbi.2014.10.007
1532-0464/

solutions to this topic such as those by Danciu et al. at Vanderbilt [23]; a report of how aggregated, de-identified data from multiple institutions can be used to address issues in pharmacosurveillance [24]; and reports addressing key methodological issues to enable data queries by overcoming issues of missing data for record linkage [25], to create transportable algorithms for clinical concept extraction across networks of disparate institutional data sets [26], and to enable advances in research queries across sites and systems in a federated fashion [27], or through the use of predefined cohorts in organized integrated data repositories [28]. As mentioned above, other reports in this special issue address methodological research that extends beyond those related directly with data re-use for research.

As in other informatics sub-domains, a core and foundational issue to be addressed involves the development of standardized methods and approaches to information integration and contextualization. To that end, Sim et al. describe a foundational ontology for clinical research, or OCRe, essentially a model for capturing the semantics underlying the scientific processes of clinical research that provides a much-needed resource to support the range of clinical research knowledge activities essential to CRI [29].

Clinical research also involves a range of steps and processes that often demand informatics solutions. Examples in this issue include approaches to: collecting and sharing trial data across study sites [30]; identifying matched controls for 'case-control' studies using EHR-based phenotyping [31]; and clustering trials with similar eligibility criteria to potentially facilitate research activities such as participant recruitment [32]. In addition, there are evolving informatics solutions designed to support the ethical imperative to accelerate research discoveries while simultaneously ensuring that research is conducted in compliance with prevailing regulatory and legal requirements [33]. A step in this direction is the included work by He et al. on a domain analysis model for e-IRB systems that helps inform their standardization, thereby enabling research by facilitating more efficient and effective research oversight [34].

Given the social nature of clinical research itself, another area that can benefit from informatics solutions involves understanding of how collaboration occurs between researchers who are often geographically distant. Using social network analysis, Bian et al. studied biomedical research collaboration networks and described researcher collaboration behavior [35]. Such findings could inform the development of tools and techniques for improving collaboration in research, particularly important in this era of "team science" and the increasing use of computer-supported collaborative work (CSCW) platforms for such endeavors.

In general, the use of information systems demands attention to the end-user's information needs, and this certainly holds true in the field of CRI. However, CRI does seem to differ from, or lag behind, domains like clinical informatics in the degree to which clinical research activities and user needs vary and tend not to be well described or understood by those creating CRI systems. To address this, Boland and colleagues describe a two-phase, mixed methods evaluation framework to inform the essential but too often overlooked or poorly-understood user-needs when iteratively developing CRI systems [36]. Such work is essential to ensuring that informatics solutions devised to address real-world research information needs are successful and anticipate or mitigate potential barriers to adoption of CRI techniques and systems.

Finally, there is a growing need to enable systematic learning and evidence-generation through the course of clinical practice, ideally by integrating research with practice, enabling evidence generating medicine activities, and creating a learning health system [19,37]. To facilitate this, we need models that inform the design and execution of practice solutions to optimize data collection and retrieval to advance research while also improving rather than impeding patient care. Work like the model-driven approach described by Curcin et al. provides methods that will help us to achieve that idealized state, advancing both research and quality of care through routine practice, all while providing a reproducible and systematic framework to organize informatics solutions for these critical areas [38].

As this special issue and the growing body of CRI methodology literature demonstrate, the efforts of CRI investigators are helping to lay a strong foundation for improvements in informatics practice as it relates to the efficient and effective conduct of clinical research. Through continued research efforts such as those highlighted in this issue, the future of CRI informatics will remain bright and the rewards will accrue to all of us who benefit from the resulting biomedical discoveries that informatics approaches will have enabled.

Acknowledgments

We would like to thank: all of the authors who submitted reports to this special issue; our colleagues who volunteered to spend countless hours reviewing and re-reviewing those reports; the JBI staff for their ongoing assistance; and Dr. Ted Shortliffe in particular for affording us in the CRI community the opportunity to document this snapshot of some of the important foundational work in our field.

References

[1] Research NDsPoC. NIH director's panel on clinical research report. Bethesda, MD: National Institutes of Health; 1997.
[2] The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. Stud Health Technol Inform 2007;129(Pt 1):330–4. PubMed PMID: 17911733. Epub 2007/10/04.eng.
[3] Kakazu KK, Cheung LW, Lynne W. The Cancer Biomedical Informatics Grid (caBIG): pioneering an expansive network of information and tools for collaborative cancer research. Hawaii Med J 2004;63(9):273–5. PubMed PMID: 15540527.
[4] Zerhouni EA. Translational and clinical science – time for a new vision. N Engl J Med 2005;353(15):1621–3. PubMed PMID: 16221788.
[5] Zerhouni EA. Clinical research at a crossroads: the NIH roadmap. J Investig Med 2006;54(4):171–3. PubMed PMID: 17152855.eng.
[6] Payne PR, Johnson SB, Starren JB, Tilson HH, Dowdy D. Breaking the translational barriers: the value of integrating biomedical informatics and translational research. J Investig Med 2005;53(4):192–200. PubMed PMID: 15974245.
[7] Sung NS, Crowley Jr WF, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA 2003;289(10):1278–87. PubMed PMID: 12633190.
[8] Oster S, Langella S, Hastings S, Ervin D, Madduri R, Phillips J, et al. caGrid 1.0: an enterprise Grid infrastructure for biomedical research. J Am Med Inform Assoc 2008;15(2):138–49. PubMed PMID: 18096909. Pubmed Central PMCID: 2274794. Epub 2007/12/22.eng.
[9] Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, et al. CaGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. Bioinformatics 2006;22(15):1910–6. PubMed PMID: 16766552.
[10] Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc: JAMIA 2009;16(3):316–27. PubMed PMID: 19261934. Pubmed Central PMCID: 2732242.
[11] Oster S, Langella S, Hastings S, Ervin D, Madduri R, Phillips J, et al. caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical Research. J Am Med Inf Assoc: JAMIA 2008;15(2):138–49. PubMed PMID: 18096909.eng.
[12] The Cancer Biomedical Informatics Grid. (caBIG): infrastructure and applications for a worldwide research community. Medinfo 2007;12(Pt 1):330–4. PubMed PMID: 17911733.
[13] Chung TK, Kukafka R, Johnson SB. Reengineering clinical research with informatics. J Investig Med 2006;54(6):327–33. PubMed PMID: 17134616.eng.
[14] Payne PR, Embi PJ, Kahn MG. Selected papers from the 2011 Summit on Clinical Research Informatics. J Biomed Inf 2011;44(Suppl 1):S54–5. PubMed PMID: 22138363.
[15] Embi PJ. Clinical research informatics: survey of recent advances and trends in a maturing field. Yearbook Med Inf 2013;8(1):178–84. PubMed PMID: 23974569.
[16] Collen MF. Clinical research databases – a historical review. J Med Syst 1990;14(6):323–44. PubMed PMID: 2132040.

[17] Gillen JE, Tse T, Ide NC, McCray AT. Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. Stud Health Technol Inf 2004;107(Pt 2):1466–70. PubMed PMID: 15361058.
[18] Holve E, Segal C, Lopez MH, Rein A, Johnson BH. The electronic data methods (EDM) forum for comparative effectiveness research (CER). Med Care 2012;50(Suppl):S7–S10. PubMed PMID: 22692262. Epub 2012/06/22.eng.
[19] Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010;2(57):57cm29. PubMed PMID: 21068440. Epub 2010/11/12.
[20] Hripcsak G, Bloomrosen M, FlatelyBrennan P, Chute CG, Cimino J, Detmer DE, et al. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. J Am Med Inf Assoc: JAMIA 2014;21(2):204–11. PubMed PMID: 24169275. Pubmed Central PMCID: 3932468.
[21] Richesson RL. An informatics framework for the standardized collection and analysis of medication data in networked research. J Biomed Inf 2014;52:4–10.
[22] Cimino J, Ayres EJ, Remennik L, Rath S, Freedman R, Beri A, et al. The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date. J Biomed Inf 2014;52:11–27.
[23] Danciu I, Cowan J, Basford M, Wang X, Saip A, Osgood S, et al. Secondary use of clinical data: the Vanderbilt approach. J Biomed Inf 2014;52:28–35.
[24] Patel VN, Kaelber DC. Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: a case study of azathioprine. J Biomed Inf 2014;52:36–42.
[25] Ong TC, Mannino MV, Schilling LM, Kahn MG. Improving record linkage performance in the presence of missing linkage data. J Biomed Inf 2014;52:43–54.
[26] Lv X, Guan Y, Deng B. Transfer learning based clinical concept extraction on data from multiple sources. J Biomed Inf 2014;52:55–64.
[27] Wyatt MC, Hendrickson RC, Ames M, Bondy J, Ranauro P, English TM, et al. Federated Aggregate Cohort Estimator (FACE): an easy to deploy, vendor neutral, multi-institutional cohort query architecture. J Biomed Inf 2013;52:156–62.
[28] Wade TD, Zelarney PT, Hum RC, McGee S, Batson D. Using patient lists to add value to integrated data repositories. J Biomed Inf 2014;52:65–70.
[29] Sim I, Tu SW, Carini S, Lehmann HP, Pollack BH, Peleg M, et al. The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research. J Biomed Inf 2014;52:71–84.
[30] Rossi E, Rosa M, Rossi L, Priori A, Marceglia S. WebBioBank: a new platform for integrating clinical forms and shared neurosignal analyses to support multi-centre studies in Parkinson's disease. J Biomed Inf 2014;52:127–39.
[31] Castro VM, Apperson WK, Gainer VS, Ananthakrishnan AN, Goodson AP, Wang TD, et al. Evaluation of matched control algorithms in EHR-based phenotyping studies: A case study of inflammatory bowel disease comorbidities. J Biomed Inf 2014;52:140–6.
[32] Hao T, Rusanov A, Boland MR, Weng C. Clustering clinical trials with similar eligibility criteria features. J Biomed Inf 2014;52:147–55.
[33] Goodman KW, Adams S, Berner ES, Embi PJ, Hsiung R, Hurdle J, et al. AMIA's code of professional and ethical conduct. J Am Med Inf Assoc: JAMIA 2013;20(1):141–3.
[34] He S, Narus SP, Facelli JC, Lau LM, Botkin JR, Hurdle J. A domain analysis model for e-IRB systems: addressing the weak link in clinical research informatics. J Biomed Inf 2014;52:85–93.
[35] Bian J, Xie M, Topaloglu U, Hudson T, Eswaran H, Hogan W. Social network analysis of biomedical research collaboration networks in a CTSA institution. J Biomed Inf 2014;52:94–104.
[36] Boland MR, Rusanov A, So Y, Lopez-Jimenez C, Busacca L, Steinman RC, et al. From expert-derived user needs to user-perceived ease of use and usefulness: a two-phase mixed-methods evaluation framework. J Biomed Inf 2014;52:105–14.
[37] Embi PJ, Payne PR. Evidence generating medicine: redefining the research–practice relationship to complete the evidence cycle. Med Care 2013;51(8 Suppl 3):S87–91.
[38] Curcin V, Woodcock T, Poots A, Majeed A, Bell D. Model-driven approach to data collection and reporting for quality improvement. J Biomed Inf 2014;52:115–26.

Peter J. Embi
Philip R.O. Payne
250 Lincoln Tower, 1800 Canon Drive, The Ohio State University, Columbus, OH 43210, USA
E-mail addresses: Peter.Embi@osumc.edu (P.J. Embi), Philip.Payne@osumc.edu (P.R.O. Payne)